OUTLIER TREATMENTS USING INTEROLATION ON MALAYSIA TOURIST ARRIVAL FORECASTING: SARIMA AND ANN APPROACHES

NORSORAYA AZURIN BINTI WAHIR

A thesis submitted in
fulfilment of the requirement for the award of the
Degree of Master of Science

Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia

JULY 2020

ACKNOWLEDGEMENT

Bismillahirahmanirahim, in the name of Allah the most merciful. Alhamdulillah, All praises for Allah, Who bestowed me the courage, spirits and power of brain to accomplish this goal. A very special thanks to my beautiful and patient supervisor Dr Maria Elena Nor for her untiring efforts, support, encouragement and leadership, and for that I will be forever grateful. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a great advisor and mentor for my master study in my life.

I would like to thank my beloved mother, grandmother and sister who always pray and deeply support for my success to complete this thesis. I also would like to express my sincere gratitude to my co-supervisors, Dr Saifullah Rusiman and Dr Khuneswari A/p P Pillay for the continuous support of my master study and research, for his patience, motivation, enthusiasm, and immense knowledge.

Last but not least, I would like to thank the rest of my thesis evaluation committee for their encouragement, insightful comments, and brilliant questions. May ALLAH shower His blessings on all of my well-wishers.

ABSTRACT

Outliers are unusual observations that appear in a piece of data that are very different from the rest of the data. The presence of an outlier may directly affect the variance, the model parameters, and the overall estimation, especially during forecasting. To obtain an accurate forecast, any outliers that are present in the data must be addressed. This research used monthly Malaysia tourist arrivals from 1998 until 2015 and an ARIMA outlier detection method to detect outliers on original data. The detected outliers were regarded as missing values then treated using interpolation method which are Linear Interpolation and Cubic Spline Interpolation methods. In this study, SARIMA model and Artificial Neural Network model were used as forecasting tools using the data before and after outlier treatment. The comparison of forecast performance between all models were calculated using MSE, MAD, MAPE and R^2 including the data before and after outlier treatment. This study found that once the outlier in the data was treated, ANN model of Cubic Spline Interpolation performs the best models compare to other models which is 95.65% using R^2 validation test. On the other hand, ANN approach outperforms SARIMA approach on both data for before and after outlier treatment which are 6.05% and 2.52%.

ABSTRAK

Nilai lampau adalah nilai ganjil yang berada dalam data dan ianya berbeza dengan nilai-nilai lain. Kehadiran nilai lampau ini memberikan kesan kepada varians, model parameter and anggaran terutamanya ketika ramalan. Hasil ramalan yang tepat dan lebih baik boleh diperoleh dengan merawat nilai terlampau yang berada dalam data sebelum proses ramalan dilakukan. Kajian ini menggunakan data kehadiran pelancong ke Malaysia dalam bulanan dari tahun 1998 hingga 2015 dan menggunakan kaedah pengesanan *ARIMA* nilai lampau untuk mengesan kehadiran nilai lampau dalam data.Nilai lampau yang dikesan dianggap sebagai nilai yang hilang kemudian dirawat menggunakan kaedah interpolasi iaitu kaedah interpolasi linear dan kaedah interpolasi spline kubik. Dalam kajian ini, model SARIMA dan model rangkaian neural digunakan sebagai model peramalan yang menggunakan data sebelum dan selepas rawatan nilai lampau. Perbandingan persembahan peramalan di antara semua model yang terlibat dinilai menggunakan *MSE, MAD, MAPE* dan ujian R^2 termasuk data sebelum dan selepas rawatan outlier. Hasil kajian ini mendapati bahawa apabila nilai lampau yang telah dirawat, model ANN daripada kaedah interpolasi spline kubik menjadi model yang terbaik berbanding model-model yang lain iaitu 95.65% menggunakan ujian keberkesanan R^2. Pada masa yang sama, kaedah rangkaian neural (ANN) memberikan persembahan yang lebih baik daripada kaedah SARIMA untuk data sebelum dan selepas rawatan nilai lampau iaitu 6.05% dan 2.52%.

\

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$\Phi_p(B)$ - Order of AR operators

s - Seasonal length (s=12 for monthly data)

$\theta_q(B)$	-	Order of MA operator
a_t	-	White noise with normal distribution $N(0, \sigma^2)$
δ	-	Constant
y_t	-	Time series data
n	-	Number of outlier
b_i	-	Bias
x_j	-	Independent variables (inputs)
$w_{i,j}$	-	Parameter (weight) form input x_j to i^{th}
y	-	Total all the output and parameter that enter the i^{th} neuron
h	-	1, 2, . . . , 5 (number of neurons)
γ_0	-	Bias for output
γ_j	-	Parameter (weight) from n_i to output
$\hat{F}(t)$	-	Final forecast value at the same original scale.
$\widehat{F1}(t)$	-	Forecast value
$f_1(x)$	-	Missing observation
$f(x_1)$	-	Ending point of gap
$f(x_0)$	-	Coordinates of starting points of gap
BJ	-	Box-Jenkins
ANN	-	Artificial Neural Network
LI	-	Linear Interpolation
CS	-	Cubic Spline
SAS	-	Statistical Analysis System

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Studies regarding outliers are common subject of research. In fact, it has a long history dating back to the earliest data analysis. An outlier is a value in a dataset that appears to behave differently than the rest of the values, where usually, the outlier value is either too small or too large. It can also be detected using certain methods. Outliers in a dataset could easily influence modelling accuracy by affecting the parameters and variables of a model. It can also affect the model estimation, leading to poor forecast accuracy and bad results. Therefore, this study detected the outlier using outlier detection method and treated the detected outlier using interpolation method as outlier treatment. Then, the outliers were regarded as missing values and the data series was subject to an outlier treatment method. Then, the forecast accuracies of the SARIMA model and the Artificial Neural Network (ANN) model using Malaysian tourist demand data were compared. The forecast accuracy of both the original and predicted data in the two models was measured to evaluate forecasting performance. In this way, the best approach could be determined. Hence, the SARIMA model was compared to ANN.

1.2 Background study

The tourism sector is the world's largest growing sector next to the foreign exchange and employment sectors. It is a crucial sector on which a lot of national economies depend, including Malaysia. In Malaysia, tourism is one of the important sectors that massively impact the country's economy. This sector also generates profound income globally because it is the sector that is primarily responsible for moving people around the world, either as tourists or immigrants. According to the Immigration Department of Malaysia, Malaysia welcomed more than 16.1 million tourists from around the globe in the first seven months of 2014, placing it 9th in the world and the 1st in Southeast Asia for the highest rate of tourist arrivals. Malaysia was given great recognition when the United Nations World Tourism Organisation (UNWTO) listed the country as the 10th most visited country in 2012. In 2017, Malaysia's tourism sector contributed around RM 82.2 billion to the country's revenue from the arrival of 25,948,459 international tourists, correlating to growth in tourist receipts by 0.1%. According Malaysia tourism website, this industry reported its half-year in positive growth of tourist arrivals, reached 13.35 million up 4.9%, while tourist receipts improved 6.8% over the same period in 2018. Malaysia also expected more tourist arrivals in 2020 for visit Malaysia's year.

In 2017, Malaysia was the second most-visited South East Asian country after Thailand with 35.3 million tourist arrivals, as stated by the Tourism Malaysia website. This makes Malaysia the fifth most visited country in Asia and the 9th best tourism country in the world, as per the Travel and Tourism Competitiveness Report 2017. As a result, Malaysia's tourism industry has become the third-largest source of foreign exchange income, owing to the consumption of goods and services by tourists. The business generated by tourist arrivals also increases the opportunity for employment and economic advancement besides advancing the transportation sector. Tourism, when combined with forecasting, could highlight the benefits of the sector and reduce outlier. Besides, government agencies need accurate data and sufficient information about forecasts for tourist demand. Also, the development distribution of infrastructures needs to be planned accurately such as the provision of comfortable accommodation, excellent transportation, and cleanliness of the environment to accommodate tourist arrivals to the country.

An accurate forecast can help investors to plan more efficiently and effectively besides securing a higher percentage of success in investment. For example, foreign investors who want to run a hotel in Malaysia or foreign businessmen who want to do business in Malaysia need to have specific, correct and suitable data to predict factors such as the weather, exchange rate, economic conditions, and security issues before deciding whether or not to invest. According to Goodwin (2008), several factors may affect tourism demand such as the impact of some events occurring in the country, e.g. war, natural disasters, diseases, as well as changes in trends and seasonality. Moreover, the importance of businesses and the macroeconomic investment by investors can also be factors that affect tourism demand.

Forecasting is a method for predicting and assuming a future event based on provided historical data in which the result will more or less resemble today rather than yesterday. This tool is very important in all sectors of the organisation since it allows the forecasting of future events including services, profit, and products. Besides, organisational analysts often use forecasting for strategic planning, finance, and accounting including budgets, production, operations, and most importantly for future sales and products or marketing. There are many types of popular time series forecasting models including using stochastic models, Artificial Neural Networks and support vector machines. The main aim of time series modelling is to collect and study the past observation of a time series to develop an appropriate model which describes the inherent structure of the series. In seasonal time series forecasting, Box and Jenkins (1976) also proposed a successful variation of ARIMA model which known as seasonal ARIMA (SARIMA). This model is popular mainly due to its flexibility to represent several varieties of time series with simplicity for optimal building process. The severe limitation of these models is pre-assumed linear form of the time series which becomes inadequate in practical situations however, various non-linear stochastic models also have been proposed but it is not so straight-forward and simple as the ARIMA model. On the other hand, for Artificial Neural Network (ANN), it's feature when it applied to time series forecasting is, its inherent capability of non-linear modelling without any presumption, about statistical distribution following by the observations.

1.3 Box-Jenkins

The Box-Jenkins approach is the most popular forecasting approach that exists in the literature, as it is one of the most powerful methods for forecasting data. This method is a systematic approach for identifying, fitting, checking, and estimating time-series models based on the inputs of the time series. The Box-Jenkins model is built using three stages: identification, estimation, and validation. These three steps are repeated several times until a satisfactory model is achieved. A time series could be in a stationary state if there is no change in trend, variance, or periodic variation. This method does not assume any pattern based on the historical data to be forecasted and is quite flexible, as it includes autoregressive and moving average terms. One of the important uses of this model to forecast a variety of data points or ranges including business data and security prices. Following the rule of thumb, at least 50 observations and preferably more than 100 observations are needed to build a proper model (Box and Tiao, 1975).

1.4 Artificial Neural Network

An Artificial Neural Network can be defined as a tool that contains several simple components that have a high correlation with each other. These components are used to process information using their system and are then fed to external inputs to give responses and results. The Artificial Neural Network consists of a large number of ‘neurons’ including simple linear and nonlinear computing elements, interconnected in complex ways that can be organised into layers. The Artificial Neural Network can be used in three ways—as a data analytic method, a model of biological nervous systems and ‘intelligence’, with real-time adaptive signal processors or controllers implanted in hardware for some field applications. In this research, the data analytic method was used. The computational system, which is a programme, starts at the first line of code, executes it, erases it, and then goes to the next by following certain instructions.

1.5 Outlier

Outlier is measurements that fit a parameterized model with certain residual which appear inconsistent with the rest of group in the sample data. This non stationary data series are more focusing on detection and the treatment of outlier. There are several types of outlier like Additive Outliers (AO), Innovation Outliers (IO), Level Shift Outliers (LSO) and Transitory Change Outliers (TCO). Outliers in time series are often influenced by interruptive events such as strikes data, economic crisis, and politics season, natural disaster like earth quake, tsunami and undetected error of typing or recording.

Outliers can be defined as abnormalities or unusual values that can be detected in dataset and may distort statistical analysis and affect the assumptions. Whereas, noise is mislabelled examples or errors in the values which can noise can be easily found in any data and it's doesn't have any pattern, unavoidable also unpredictable. Noise generally consists of residuals and errors. Residuals are the differences between the observed and predicted values. Data points which have large residuals is called an outlier. Then, errors may include measurement errors and sampling errors and most errors are unavoidable while unless systematic errors.

There are several methods to treat the outliers which are univariate method, multivariate method and *Minkowski* error. *Minkowski* error is the method to reduce the contribution of potential outliers in the training process whereas the univariate and multivariate methods are looking for data points with extreme values. However, the focus in this research is to detect and handle the outlier in univariate. The implication of the data series before and after handling the outlier were made to analyse the effects of outlier in the data. Although there are many popular outlier treatments like robust statistics, outlier removal clustering, trimming function, Winsorized estimators and bootstrapping, interpolation is one of the great methods in dealing with missing values since in this study, the outlier was regarded with missing value.

1.6 Problem statement

Data that contains an outlier acts differently than others and have significant possibilities of interrupting the analysis, making the data estimation inaccurate.

Barnett and Lewis (1994) posed a great discussion about an identified outlier that is illegitimately present in the data to which one of the best methods is to discard it. When the outlier is present as a legitimate part of the data, the issue becomes more serious. However, Orr, Sackett, and DuBois (1991) proved that there is disagreement among researchers on whether deleting outliers in the data without treating them as outliers are the best course. The removal of outliers may produce undesirable outcomes when the outliers are illegitimate since it may affect size measures in studies and variances. According to Pigott (2001), many researchers face the same problems in handling this issue. The process of collecting data had to be redone and started over again to avoid outliers. This process extended the time it took to obtain the results and wasted money and energy. The literature abounds with studies that support removing outliers versus those that propose replacing outliers with new values which are Winsorization method.

Several previous literatures like Rose, Ismail and Rosli (2019) and Barrow and Kounrentzes (2018) are focusing on detection of outlier instead of dealing with outlier when using time series data. Besides, there are also limited studies that have applied outlier treatments in a time series especially using Malaysia's data. Most of the previous works are focusing on the best tool and method to detect the outlier in data. The literature review also shows limited of studies that have investigated the effect of outliers and implication on the forecast performance of time series data. Thus, this research proposed an alternative method which are linear interpolation and cubic spline interpolation to treat outliers in time series data. Hence, the forecast performances were compared before and after outlier treatment. The interpolation method easier to access by anyone and its does not requires a more complicated computation method like robust regression or other methods. This method assumed that the huge fluctuations (outlier) does not occur within the data and the desirable time span available should in uniform one which suitable with time series data.

1.7 Objectives

There are few of objective in these studies which are:

(i) To evaluate the outlier in the time series data using iterative outlier detection method.

(ii) To propose the Linear and Cubic Spline Interpolation as outlier treatment.

(iii) To investigate the forecast performance between SARIMA method and Artificial Neural Network method before and after outlier treatment.

1.8 Scope of study

The effect of SARIMA and Artificial Neural Network forecasting performance before and after the outlier treatment in time series data is the main focus of this study. The monthly data of Malaysian tourist arrivals was obtained from the official website of the Department of Statistics Malaysia from January 1998 to December 2015. This study only focused on detecting outlier using iterative outlier detection method and at the same time, handling outliers in the data, so two outlier interpolation models were considered: Linear and Cubic Spline. To determine the best model for treating the outliers, the forecast accuracy of each model was compared in terms of mean squared error (MSE), mean absolute deviation (MAD), and mean absolute percentage error (MAPE) including model validation test using R-squared values. In this study, the detected outliers were using S-PLUS software version 9.3 whereas the outlier treatment was done using SRS1 Spline software. This software is to simplify the process of treating the outliers since its' are easily to access in any market and familiar which can be applied in Microsoft Excel for everyone. Moreover, this method is much more straightforward to compute in computational ways compared to the other software like R, Phyton and C++ software.

1.9 Significant of study

This research will benefit researchers, investors, planners, and analysts especially those working in organisations related to forecasting time series data such as Department of Education, Statistics, Finance, Health and others. The most past studies, detecting outliers was reportedly easier when sophisticated software was used, but difficult cases still required the specific handling and treatment of outliers. The current study proposes treating outliers in time series data using the interpolation method and examining the performance of the forecasting model before and after the treatment. Even, there are some previous studies are using Winsorization method to treat the

outlier in the data, none of the studies on time series data are specifically using interpolation method to treat outliers; this study is the first to do so.

1.10 Summary

This chapter explained the outliers and their effects in relationship with the time series dataset. This study aims to determine the effect of outliers on the forecast performance of SARIMA and the Artificial Neural Network (ANN). Moreover, this study also compared the forecast accuracy of these models before and after outlier treatment. The outliers were detected using S-PLUS software and two interpolation methods (Linear and Cubic spline) were tested. Hence, the best performance in forecasting time series data determined. The data were analysed using Minitab and SAS software. The rest of the thesis is outlined as follows: Chapter 2 provides a brief review of the methods and issues of Artificial Neural Network and SARIMA time series interpolation methods. Chapter 3 deals with the research design procedure for replacing outliers with new values and the effect of both outlier treatment methods on the data. Chapter 4 presents the empirical findings and discussions of the proposed methods. Finally, conclusions and some recommendations are presented in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter is divided into seven sections, namely using Box-Jenkins for time series forecasting, using the Artificial Neural Network for time series forecasting, outliers in time series data, outlier treatment in time series data, the effect of outliers in a time series, the interpolation of outliers in a time series, and the forecasting of tourism data.

2.2 Box-Jenkins for time series forecasting

The Box-Jenkins approach or generally known as the ARIMA model is one of the traditional methods for analysing time series data developed by Box and Jenkins (1976). Box and Jenkins (1976) first introduced the time-series ARIMA model with three parts, which are auto regression AR (p), moving averages MA (q), and differencing (d). To strip off the integration (I) of the series (d), which forms ARIMA (p, d, q), the data must be in a stationary state (Lin, Chen & Lee, 2011). Zhang and Qi (2005) stated that differencing must be applied to achieve stationarity. The most appropriate parameters (p and q) for the time series model, which show the system behaviour, are determined via the evaluation of the autocorrelation function (ACF) and partial autocorrelation function (PACF) (Abudu, King & Bawazir 2010; Abudu, King, & Sheng 2012; Sabzi, King, & Abudu 2017).

Shrivastav (2012) used the Box-Jenkins model to predict crime incidents based on previous crime data in Gujarat State about the counterfeiting of currency. This

forecast helped to strengthen valuable sources for law agencies and helped increase the police performance via strategic deployment efforts. It also afforded more advantages to making an efficient investigation. In the same year, Garrett (2012) forecasted future disease based on historical local public health records. The research indicated that the Box-Jenkins approach produced 75% forecasting accuracy. In recent years, the ARIMA and SARIMA models have gained much popularity in the tourism industry, particularly among the decision-makers (Sufahani, Ismail, & Muhammad (2013). The ARIMA model can also be used in econometrics. Theresa (2013) used the Box-Jenkins method to identify a suitable time series model based on the sample autocorrelation function (ACF), the partial autocorrelation (PACF), and the theoretical autocorrelation function. Although the ARIMA approach is popular and has been used in many studies, this approach is only suitable when the time series is assumed to be stationary (Singh and Mishra, 2015).

Mondal, Shit & Goswami (2014) investigated the effectiveness of an ARIMA time series model for Indian stock price forecasting from different sectors. The result showed that the ARIMA model had 85% more accuracy in predicting the stock price compared to other forecasting models, even when data from different sectors were used. Dritsaki (2016) forecasted the real GDP rate using the ARIMA model and noted an increasing trend. Wang *et al.* (2015) attempted to improve the accuracy of forecasting annual runoff time series by combining the ARIMA model with ensemble empirical mode decomposition (EEMD). The result showed that EEMD could enhance forecasting accuracy, significantly helping to improve the ARIMA model in forecasting the annual runoff time series data. Unhapipat (2018) use Box-Jenkins (ARIMA) method which is ARIMA (0,0,0) x (1,1,0) to analyse the international tourist visit to Bumthang with 91% accuracy. These forecasted results were used as a tool to predict the future challenges and to bring further development in tourism sector.

Ismail (2019) also applied ARIMA model to predict and evaluate the short-term forecasting of the Arab and Foreign tourist in Eygpt based on time series data. Ahire, Fernandes & Teixeiral (2020) analyse the growth trends in medical tourism in India and to forecast the medical tourist arrivals using ARIMA method for trend projection. The analysis shows this sector denotes a significant growth and a great potential to earn valuable foreign exchanges through it. Singh & Mishra (2015) forecasted the prices of Groundnut oil in Mumbai and showed that ANN performed better than the ARIMA model in time series forecasting.

REFERENCES

Ahelegbey, D. F. (2015). The econometrics of networks: A Review. *Working paper, 13*.

Abudu, S., King, J. P., & Bawazir, A. S. (2010). Forecasting monthly stream flow of spring-summer run off season in Rio Grande headwaters basin using stochast ic hybrid modelling approach. *Journal of Hydrologic Engineering*, 16(4), 384-390.

Abudu, S., King, J. P., & Sheng, Z. (2012). Comparison of the performance of statistical models in forecasting monthly total dissolved solids in the Rio Grande 1. *JAWRA Journal of the American Water Resources Association*, 48(1), 10-23.

Acimovic, J., Erize, F., Hu, K., Thomas, D. J., & Mieghem, J. A. V. (2018). *Product life cycle data set: raw and cleaned data of weekly orders for personal computers*. Manufacturing and Service Operations Management.

Abraham, B., and Box, G. E. (1979). Bayesian analysis of some outlier problems in time series. *Biometrika*, *66*(2), 229-236.

Ahmed, R., Vveinhardt, J., Ahmad, N., & Streimikiene, D. (2017). Karachi inter-bank offered rate (KIBOR) forecasting: Box-Jenkins (ARIMA) testing approach. *E&M Economics and Management*, 20(2), 188-198.

Ahmar, A. S., Guritno, S., Rahman, A., Minggi, I., Tiro, M. A., Aidid, M. K, & Ahmar, A. A. (2018). Modeling data containing outliers using arima Additive outlier (ARIMA-AO). In *Journal of Physics: Conference Series* (Vol. 954, No. 1, p. 012010). IOP Publishing.

Arumugam, P., and Saranya, R. (2018). Outlier detection and missing value in seasonal ARIMA model using rainfall data. *Materials Today: Proceedings*, 5(1), 1791-1799.

Alvarez, E., Garcıa-Fernández, R. M., Blanco-Encomienda, F. J., & Munoz, J. F. (2014). The effect of outliers on the economic and social survey on income and living conditions. *World Acad. Sci., Eng. Technol., Int. J. Soc., Behav., Educ., Econ., Bus. Ind. Eng*, 8.3276-3280.

Alice, C. & Bovas, A., (1989). "Comparison of parameter estimation methods in time series outliers: a simulation study" in *ASA Proceedings, Business and Economic Statistics Section*.Vol. 4.83-92.

Byers, J. W., Popova, I., and Simkins, B. J. (2018). The impact of outliers on computing conditional risk measures for crude oil and natural gas commodity futures prices

Barnett, V., & Lewis, T. (1994). Wiley series in probability and mathematical statistics: Applied probability and statistics. *Outliers in Statistical Data*. 346-372

Battaglias,F. (2006). "On outlier detection in multivariate time series". Unpublished paper presented, *Seminar in the University of Liverpool, United Kingdom.*

Box, G. E. P. and Jenkins G. M. (1976). Time series analysis. *Forecasting and control. Holden-Day*. San Francisco.

Baek, J., and Cho, S. (2003). Bankruptcy prediction for credit risk using an auto-associative neural network in Korean firms. In *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference*. 25-29.

Barnett, V., and Lewis, T. (1978). *Outliers in statistics*. New York: Wiley

Bustos, O. H., and Yohai, V. J. (1986). Robust estimates for ARMA models. *Journal of the American Statistical Association*, *81*(393). 155-168.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1978). *Time series analysis: forecasting and control*. Princeton-Hall International.

Chan, M. C., Wong, C. C., and Lam, C. C. (2000). Financial time series forecasting by Neural Network using conjugate gradient learning algorithm and multiple linear regression weight initialization. *Computing in Economics and Finance*, 61.326-34.

Chang, I., and Tiao, G. C. (1983). Effect of exogenous interventions on the Estimation of time series parameters. In *Proceedings of the American Statistical Association, Business and Economics Statistics Sectio*,532-537.

Chen, C., and Liu, L. M. (1993). Joint estimation of model parameters and outlier effect in time series. *Journal of the American Statistical Association*, *88*.421. 284-297.

Cousineau, D., and Chartier, S. (2010). Outlier detection and treatment: a review. *International Journal of Psychological Research*, *3*(1).58-67.

Chang, I., and Tiao, G. C. (1983). Effect of exogenous interventions on the estimation of time series parameters. In *Proceedings of the American Statistical Association, Business and Economics Statistics Section*. 532-537.

Crouch, G. I. (1994). The Study of international tourism demand: A survey of practice. *Journal of Travel Research*, *32*(4). 41-55.

Chin Foon Khoo, Sharifah Sakinah Syed Ahmad, and Zuraini Othman. (2008). *Numerical Methods*. Prentice Hal Pearson, Petaling Jaya

Calantone, R. J., Di Benedetto, A., and Bojanic, D. C. (1988). Multimethod forecasts for tourism analysis. *Annals of Tourism Research*, *15*(3), 387- 406.

Chik, Z. (2002). The effect of outliers on the performance of order selection criteria for short time series. *Journal of Applied Sciences, 2, 912-915.*

Dritsaki, C. (2016). Forecast of SARIMA Models: An application to unemployment rates of Greece. *American Journal of Applied Mathematics and Statistics*, *4*(5).136-148.

Durbin, J. (1979) Comment to Kleiner, Martin and Thomson: Robust estimation of power spectra. *Journal of the Royal Statistical Society*. Series B, 41. 313- 351.

Daszykowski, M., Kaczmarek, K., Vander Heyden, Y., & Walczak, B. (2007). Robust statistics in data analysis-a review basic concepts. *Chemometrics And Intelligent Laboratory Systems*, 85. 203-219.

Denby, L., and Martin, R. D. (1979). Robust estimation of the first-order autoregressive parameter. *Journal of the American Statistical Association*, *74*(365).140-146.

Dang, X., Serfling, R., and Zhou, W. (2009). Influence functions of some depth functions and application to Depth-Weighted L-Statistics. *Journal of Nonparametric Statistics*, *21*(1).49-66.

Donders, A. R. T., Van Der Heijden, G. J., Stijnen, T., and Moons, K. G. (2006). A Gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*. *59*(10).1087-1091.

Domonkos, P., and Coll, J. (2017). Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests. *International Journal of Climatology*, *37*(4).1910-1921.

Deutsch, S. J., Richards, J. E., and Swain, J. J. (1990). Effects of a single outlier on ARMA identification. *Communications in statistics-Theory and Methods*, *19*(6).2207-2227.

Elliott, M. R., and Stettler, N. (2007). Using a mixture model for multiple imputation in the presence of outliers: The 'Healthy for Life' Project. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *56*(1).63-78.

Fishwick, P. A. (1989). Qualitative Methodology in Simulation Model Engineering. *Simulation*, *52*(3).95-101.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*. 350-363.

Fröhlich, M. (2018). Nowcasting Austrian Short-Term Statistics. *Journal of Official Statistics*, *34*(2).503-522.

Faraway, J., and Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the airline data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(2).231-250.

Goodwin, P. (2008). A Quick Tour of Tourism Forecasting. *Foresight*, *10*.35-37.

Garrett, L. C. (2012). Using Box-Jenkins Modelling Techniques to Forecast Future Disease Burden and Identify Disease Aberrations. *Public Health Surveillance Report*.

Gheyas, I. A., and Smith, L. S. (2009). A Neural Network Approach to Time Series Forecasting. *In Proceedings of the World Congress on Engineering*, 2.1-3.

Gomez, V., & Maravall, A. (2001). Automatic modeling methods for univariate series. *A course in time series analysis*, 171-201.

Gomez, L. A. B., and Cappello, F. (2015). Detecting and correcting data corruption in stencil applications through multivariate interpolation. In *Cluster Computing (CLUSTER), 2015 IEEE International Conference on*,595-602.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, *11*(1).1-21

Guo, J., Huang, W., and Williams, B. M. (2015). Real time traffic flow outlier detection using short-term traffic conditional variance prediction. *Transportation Research Part C: Emerging Technologies*, *50*. 160-172.

Gnanadesikan, R., (1977). Methods for the statistical data analysis of multivariate observations. John Wiley and Sons, New York,311.

Hassan, S., & Othman, Z. (2018). Forecasting on the long term sustainability of the employees provident fund in Malaysia via the Box-Jenkins' ARIMA model. *Business and Economic Horizons (BEH)*, *14*(1232-2019-736).43.

Hansen, J. V., and Nelson, R. D. (2003). Forecasting and recombining time-series components by using neural networks. *Journal of the Operational Research Society*, *54*(3).307-317.

Han, H., and Qiao, J. (2010). A self-organizing fuzzy neural network based on a growing-and-pruning algorithm. *IEEE Transactions on Fuzzy Systems*, 18(6).1129-1143.

Hawkins, D. M. (1980). *Identification of outliers* (Vol. 11). London: Chapman and Hall.

Huynh, D. Q., Hartley, R., and Heyden, A. (2003). Outlier correction in image sequences for the affine camera. In *ICCV* .585-590.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., and Razbash, S. (2019). Package 'forecast'. *Online] https://cran. R-project. org/web/packages/forecast/forecast. pdf*.

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

Harrison, P. J., and Stevens, C. F. (1976). *Bayesian forecasting (with discussion)*. JR Statist. Soc. B., Harrison 20538J. R. *Statist. Soc*. 38. 205-247.

Ismail, S. (2008). Accommodation of outliers in time series data: A case study *Asian Journal of Mathematics and Statistics*. 1 (1) .24-33.

Ishikawa A, Endo S, and Shiratori T. (2010). Treatment of outliers in business surveys: the case of short-term economic surveys of enterprises in Japan. .*Bank in Japan Working Paper Series,*10.(E-8).

Judd, C. M., and McClelland, G. H., (1989). Data Analysis: A Model Comparison Approach. *Harcourt Brace Jovanovich*. San Diego, CA.

Kondratenko, V. V., and Kuperin, Y. A. (2003). Using Recurrent Neural Networks to forecasting of forex. *Arxiv Preprint Cond-Mat/0304469*.

Kaiser, R., & Maravall Herrero, A. (1999). Seasonal outliers in time series. Documentos de trabajo/Banco de España, 9915.

Khamis, A., Ismail, Z., Haron, K., and Mohammed, A. T. (2005). The effects of outliers data on Neural Network performance. *Journal of Applied Sciences*, *5*(8):1394-1398.

Ko, K. M., Chen, J. F., Nguyen, T. L., Hsu, B. M., and Shu, M. H. (2014). Forecasting inbound tourism demand in Thailand with grey model. *WSEAS Transactions on Mathematics*, *13*.96-104.

Lin, C. J., Chen, H. F., and Lee, T. S. (2011). Forecasting tourism demand using time series, Artificial Neural Networks and multivariate adaptive regression splines: evidence from taiwan. *International Journal of Business Administration*, *2*(2).14.

López-de-Lacalle, J. (2016). Tsoutliers R package for detection of outliers in time Series. *CRAN, R Package*.

Leskinen, E. (1983). *Notes on effects of a disturbed observation in estimation of autocovariances and autocorrelations,* Reports on Statistics, Department of Statistics, University of Jyväskylä.

Li, G., Song, H., and Witt, S. F. (2006). Time varying parameter and fixed parameter linear aids: an application to tourism demand forecasting. *International Journal of Forecasting*, *22*(1).57-71.

Mondal, P., Shit, L., and Goswami, S. (2014). Study of effectiveness of time series modelling (ARIMA) in forecasting stock prices. *International Journal of Computer Science, Engineering and Applications*, *4*(2).13.

Maity, B., and Chatterjee, B. (2012). Forecasting GDP Growth Rates of India: An Empirical Study. *International Journal of Economics and Management Sciences*, *1*(9).52-58.

Martin, R. D. (1980). *Robust estimation of autoregressive models in directions in time series*, eds. D. R. Brillinger and G. C. Tiao, Hayward, CA: Institute of Mathematical Statistics, 228-254.

Martin, R. D., and Yohai, V. J. (1986). Influence Functional for Time Series. *The Annals of Statistics*. 781-818

Muhammad, H. L., Maria, E. N., Hossain, J. S., Nur, H., and Nur, A. (2012). Fuzzy time series: An application to tourism demand forecasting. *American Journal of Applied Sciences*, *9*(1).132-140.

Nare, H., Maposa, D., and Lesaoana, M. (2012). A method for detection and Correction of outliers in time series data. *African Journal of Business Management*, *6*(22).6631-6639.

Nanthakumar, L., Ibrahim, Y., and Harun, M. (2007). *Tourism Development Policy, Strategic Alliances and Impact of Consumer Price Index on Tourist Arrivals*: The case of Malaysia.

Orr, J. M., Sackett, P. R., & Dubois, C. L. (1991). Outlier detection and treatment in i/o psychology: A survey of researcher beliefs and an empirical Illustration. *Personnel Psychology*, *44*(3).473-486.

Pigott, T. D., (2001). A Review of methods for missing data in *educational research and evaluation*. Vol. 7. No. 4 pp. 353-383.

Park, Y. R., Murray, T. J., and Chen, C. (1996). Predicting sun spots using a layered perceptron Neural Network. *IEEE Transactions on Neural Networks*.*7*(2), 501-505.

Palmer, A., Montano, J. J., and Sesé, A. (2006). Designing an Artificial Neural Network for forecasting tourism time series. *Tourism Management*, *27*(5). 781-790.

Philip, A. A., Taofiki, A. A., and Bidemi, A. A. (2011). Artificial Neural Network model for forecasting foreign exchange rate. *World of Computer Science and Information Technology Journal*, *1*(3).110-118.

Pan, Z., Liu, P., Gao, S., Feng, M., and Zhang, Y. (2018). Evaluation of flood season segmentation using seasonal exceedance probability measurement after outlier identification in the three Gorges reservoir. *Stochastic Environmental Research and Risk Assessment*, 1-14.

Petrevska, B. (2017). Predicting tourism demand by ARIMA models. *Economic research-Ekonomska istraživanja*, *30*(1).939-950.

Proietti, T., and Lütkepohl, H. (2013). Does the Box–Cox Transformation help in forecasting macroeconomic time series? *International Journal of Forecasting*, *29*(1). 88-99.

Queenan, C. C., Ferguson, M., Higbie, J., and Kapoor, R. (2007). A comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management*, *16*(6).729-746.

Roberts, S.W. (2000). Control chart tests based on geometric moving averages. *Technometrics*. 42(1). 97–101.

Sarle, W. S., (1994). *Neural Network Implementation in SAS® Software*.

Sabzi, H. Z., King, J. P., and Abudu, S. (2017). Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: A case study. *Expert Systems with Applications*, *83*.145-163.

Shrivastav, A. K. (2012). Applicability of Box Jenkins ARIMA Model in crime forecasting: A case study of counterfeiting in Gujarat state. *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, *1*(4).494.

Sufahani, S. F., Ismail, Z., and Muhammad, M. (2013). *An analysis of international tourist behavior towards tourism sector in Kelantan.*

Singh, A., and Mishra, G. C. (2015). Application of Box-Jenkins Method and Artificial Neural Network procedure for time series forecasting of prices. *Statistics in Transition new series*, *1*(16).83-96.

Sharda, R., and Patil, R. B. (1990). Neural Networks as forecasting experts: an empirical test. in *Proceedings of the International Joint Conference on Neural Networks IEEE*, Vol. 2. 491-494.

Shi, K. L., Chan, T. F., Wong, Y. K., and Ho, S. L. (2000). Direct self -control of induction motor based on Neural Network. In *Industry Applications Conference, 2000. Conference Record of the 2000 IEEE*, 3. 1380-1387.

Sulaiman, J., and Wahab, S. H. (2018). Heavy rainfall forecasting model using Artificial Neural Network for flood prone area. In *IT Convergence and Security 2017*. Springer, Singapore.68-76.

Smith, A. F. M., and West, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman Filter. *Biometrics*. 867-878.

Staal, O. M., Saelid, S., Fougner, A. L., and Stavdahl, Ø. (2018). Kalman Smoothing for objective and automatic preprocessing of glucose data.

Song, H., Li, G., Witt, S. F., and Fei, B. (2010). Tourism demand modelling and forecasting: How should demand be measured? *Tourism Economics*, 16(1). 63-81.

Song, H., and Witt, S. F. (2000). *Tourism demand modelling and forecasting: modern econometric approaches*. Routledge.

Song, H., and Li, G. (2008). Tourism demand modelling and forecasting - A review of recent research. *Tourism Management*, *29*(2).203-220.

Safi, S. K. (2016). A comparison of artificial neural network and time series models for forecasting GDP in Palestine. *American Journal of Theoretical and Applied Statistics*,5(2).58-63.

Theresa, H. D. N. (2013). The Box-Jenkins methodology for time series models. *SAS global forum.* Warner Bros. Entertainment Group, Burbank, California.

Tang, Z., and Fishwick, P. A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing*, *5*(4). 374-385.

Tang, Z., De Almeida, C., and Fishwick, P. A. (1991). Time series forecasting using Neural Networks vs. Box-Jenkins methodology. *Simulation*, *57*(5). 303-310

Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics*. Allyn and Bacon/Pearson Education.

Tsay, R. S. (1989). Identifying multivariate time series models. *Journal of Time Series Analysis*, *10*(4).357-372.

Trívez, F. J. (1995). Level shifts, temporary changes and forecasting. *Journal of Forecasting*, *14*(6). 543-550.

Van Amerom, J. F., Lloyd, D. F., Price, A. N., Kuklisova Murgasova, M., Aljabar, P., Malik, S. J., and Hajnal, J. V. (2018). Fetal cardiac cine imaging using highly accelerated dynamic MRI with retrospective motion correction and outlier rejection. *Magnetic Resonance in Medicine*. *79*(1). 327-338.

Virili, F., and Freisleben, B. (2000). Non-stationarity and data pre-processing for Neural Network predictions of an economic time series. In *IJCNN, IEEE*, 5129.

Wang, H., Gao, Q., Feng, L., Wei, R., and Wang, J. (2015). Proper orthogonal decomposition based outlier correction for PIV data. *Experiments in Fluids*, *56*(2).43.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, *80*(389).73-83.

Wang, H., Gao, Q., Feng, L., Wei, R., & Wang, J. (2015). Proper orthogonal decomposition based outlier correction for PIV data. *Experiments in Fluids*, *56*(2), 43.

Yüksel, S. (2007). an integrated forecasting approach to hotel demand. *Mathematical and Computer Modelling*, *46*(7-8).1063-1070.

Zhang, G. P., and Qi, M. (2005). Neural Network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2). 501-514.

Zakai, M. (2014). A time series modeling on GDP of Pakistan. *Journal of Contemporary Issues in Business Research*, *3*(4).200-210.

Zellner, A. (1981). Philosophy and objectives of econometrics in Currie D, Nobay R, Peel D (1981). Macroeconomic Analysis: Essays in Macroeconomics and Economics. *Croom Helm*, London. 24 - 34.

Zhuang, Y., and Chen, L. (2006). In-network outlier cleaning for data collection in sensor networks. In *Clean DB*